

AD \_\_\_\_\_

Award Number: W81XWH-12-1-0480

TITLE: Molecular Characterization of Indolent Prostate Cancer

PRINCIPAL INVESTIGATOR: Jun Luo, Ph.D.

CONTRACTING ORGANIZATION: Johns Hopkins University  
Baltimore, MD 21218-2680

REPORT DATE: December 2016

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
1. REPORT DATE December 2016		2. REPORT TYPE Final Report		3. DATES COVERED 15 Sep 2012 - 14 Sep 2016	
4. TITLE AND SUBTITLE Molecular Characterization of Indolent Prostate Cancer				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-12-1-0480	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Jun Luo, Ph.D.  E-Mail: jluo1@jhmi.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Johns Hopkins University, Baltimore, MD 21218-2680				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Indolent prostate cancers that pose very low risk to aged men occur frequently and may be detected at biopsy, leading to the contemporary problem of prostate cancer over-diagnosis and over-treatment. The objective of the project is to define and characterize indolent prostate cancer using genomic approaches in the clinically relevant context of a cohort meeting the entry criteria for active surveillance. During the funding period, we collaborated with urologists and pathologists involved in the active surveillance program, and focused on evaluation of technical feasibility, acquisition of specimen resources, and optimization of the workflow. While this exploratory work established that high quality nucleic acid molecules can be extracted from select tumor lesions, comprehensive analysis of a large cohort of men diagnosed with very low risk prostate cancer may require further technical optimization as well as optimization of clinical workflow to ensure compatibility with high-throughput studies of small pathological lesions.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT  UU	18. NUMBER OF PAGES  16	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

## Table of Contents

	<u>Page</u>
Introduction.....	1
Body.....	1-7
Key Research Accomplishments.....	7-8
Reportable Outcomes.....	8
Conclusion.....	8
References.....	8-9
Appendices.....	9
Supporting Data.....	9-16

## **Introduction**

Indolent prostate cancers pose very low risk to aged men. However, these cancers are known to occur frequently, may be detected at biopsy following serum prostate-specific antigen (PSA) test, and treated aggressively following diagnosis, leading to the contemporary problem of prostate cancer over-diagnosis and over-treatment. One way to address this problem is active surveillance. However, patient selection for active surveillance is mainly based on clinical variables. On the basis of the general concept that progressive acquisition of genomic alterations, both genetic and epigenetic, is a defining feature of all human cancers at different stages of disease progression, we hypothesized that RNA and DNA alterations characteristic of indolent prostate tumors may be different from those in clinically significant prostate cancer, and proposed a series of exploratory studies to evaluate the molecular signature of indolent prostate cancers. However, Molecular analysis of small volume, very low risk, indolent prostate tumors has not been systemically performed using genome-wide approaches mainly due to a number of technical constraints. The primary purpose of the project is to characterize indolent prostate cancer using genomic approaches in the context of a cohort of men predicted to harbor very low-risk prostate cancer at the time of biopsy detection and thus meeting the entry criteria for active surveillance. The scope of the proposed research is: 1) to define the expression signature of indolent prostate cancer by genome-wide expression analysis comparing tissue lesions from very low risk prostate cancer versus high risk prostate cancer defined by pathological outcome measures in men meeting the entry criteria for active surveillance but opting for immediate surgical treatment; 2) to develop a refined signature using biopsy specimens from an active surveillance cohort; and 3) to differentiate indolent prostate cancer from clinically significant prostate cancer using advanced deep-sequencing technologies for both DNA copy number of methylation analysis.

## **Body**

### **Findings resulting from Task 1: To define indolent human prostate cancer by genome-wide expression analysis comparing tissue lesions from RRP-confirmed very low-risk prostate cancer versus higher-risk prostate cancer (Months 1-24).**

**Summary:** During year 1 of the project period, we completed two critical project milestones associated with Task 1. First we identified men meeting the active surveillance criteria but that opted for radical retropubic prostatectomy (RRP) treatment, making it possible to perform studies utilizing these pathological specimens representing indolent prostate tumors confirmed by pathological findings. Second we performed preliminary studies using such specimens, establishing that small-volume tumors present in FFPE sections are compatible with genome-wide RNA analysis. This accomplishment was reported in our 2013 annual progress report. During year 2 of the project period, we focused on additional technical evaluation of genome-wide approaches utilized for comparison of low-risk and high-risk prostate cancer tissues collected in the standard

clinical setting involving formalin-fixation and paraffin embedding (FFPE) of the specimens. On the basis of the findings and the technical trend that was not foreseen at the time of our original grant application, we proposed slightly revised approaches employing RNA sequencing. Progresses made during this period as well as our revised approaches were reported in our 2014 annual progress report. During year 3 of the project period, we focused our efforts in accrual of suitable clinical specimens for this revised study. Following review of our progress on the tasks outlined in SOW, we communicated to CDMRP our intention to request EWOFF. On October 15<sup>th</sup>, 2015, this project was officially approved for an EWOFF of an additional 12 months. During the EWOFF period, we invested our efforts in careful evaluation of RNA extracted from very small volume tumor present in a single pathological slide. We evaluated a total of 70 cases (details below). Unexpectedly, the vast majority of indolent prostate cancer cases present very small lesions, compromising RNA yield and quality. Both RNA yield and quality also varied among the samples. Following these extensive efforts, we determined that RNA sequencing is unlikely to succeed given the current technology available to us, due to the combination of low RNA yield and low RNA quality from these pathological specimens

**Supporting data:** Supporting data Figures and Tables can be found in the Appendix of this Final Report.

#### **Year 1:**

Identification of RRP-confirmed indolent prostate tumors. We performed a survey of surgical prostate cases from men meeting the active surveillance criteria in our institution (i.e., the Epstein criteria) over a 3-month period. The goal was to determine whether it was feasible to acquire sufficient number of recently processed FFPE cases that are from men operated for prostate cancer despite meeting the entry criteria for active surveillance. This survey was critical because our previous studies (1, 2) have shown that high-fidelity genomic data can be obtained from these recently processed FFPE specimens. Over a 3-month period, we identified 44 RRP cases fulfilling the pathological criteria for active surveillance. Of these, 21 were organ confined Gleason score 6 consistent with the definition of clinically insignificant (i.e., indolent) prostate tumors, while 23 were upgraded. On the basis of these findings, we were assured that the required number of indolent prostate cancer cases operated within a defined time window may be acquired for genome-wide expression studies.

High-quality nucleic acid samples obtained from tissues with indolent prostate cancer. In previous studies, we tackled a number of technical variables relevant to genome-wide expression analysis of formalin fixed paraffin embedded (FFPE) prostate tissue specimens (1, 2). However, our optimized technical procedures had not been tested in small tumors present in FFPE sections from men qualified for active surveillance. We consider a further optimized workflow using target specimens (from men who qualified for active surveillance but opted for surgery) an essential step toward the generation of high-fidelity genomic data. We performed laser-capture microdissection (LCM) (Figure 1, Supporting Data) and downstream RNA extraction (Figure 2, Supporting Data). We show that good

quality of RNA sufficient for the proposed studies can be consistently extracted from such cases (Figure 2, Supporting Data).

## **Year 2:**

RNA-Seq approach for the comparison of low-risk and high-risk prostate tumors. During this project period, the general research field of genome profiling underwent some drastic changes. Specifically, RNA sequencing is replacing the traditional expression microarray as the standard methodology for analysis of the entire transcriptome. It is important to adapt to this technical trend. Nevertheless RNA-Seq in paraffin-embedded specimens needs to be fully evaluated under laboratory-specific conditions with full implementation of quality control measures to ensure data validity. We noted that additional technical advances have been made that are relevant to RNA-Seq using limited amount of FFPE materials. For example, in studies comparing different RNA-Seq library preparation methods using degraded and/or low-input RNA samples (3, 4), a number of key RNA-Seq technical metrics were evaluated, demonstrating the overall feasibility of achieving 1) efficient rRNA depletion (down to 0.1% of reads aligned to rRNA genes) (3, 4), an essential step in RNA-Seq of FFPE RNA; 2) genome alignment of reads at levels equivalent to RNA-Seq reads from gold-standard high-quality mRNA from fresh frozen samples (3, 4); 3) High sensitivity in transcript detection (3, 4); 4) Acceptable % of exon coverage (greater than 40% of reads mapping to exons) (3, 4); 5) Uniform transcript coverage (3, 4); 6) High concordance in transcript quantification between FFPE RNA-Seq and expression microarrays of fresh frozen tissues(3, 4), at a level similar to the comparison between different expression microarray platforms.

It is in the context of these latest technical advances that we performed preliminary studies evaluating RNA-Seq using FFPE specimens that are used in the comparison of low-risk and high-risk prostate cancer. We presented summary data derived from two cases (59642 and 59643). We used 3 different starting amounts (200pg, 2ng 10ng rRNA depleted RNA) of FFPE RNA to make sequencing libraries. We used the rRNA-depletion protocol with Clontech RoboGone-Mammalian kit(cat#634846 Clontech , USA). After rRNA depletion, cDNA synthesis was made with SMARTer Universal Low Input RNA kit from Clontech. This kit starts with low amount of input RNA then a modified N6 primer (the SMART N6 CDS primer) for first-strand synthesis. The SMARTScribe Reverse Transcriptase enables template switching and extension to produce the complementary DNA strand. After cDNA amplification, final amplified cDNA is digested with RsaI to remove the SMART adapter. Following the Low Input Library Prep kit, FFPE RNA-Seq library was generated. We quantified final libraries with Agilent bioanalyzer and measured with Invitrogen Qubit. All 6 RNA samples were added different indexes to be pooled together for one lane of 50bp single read sequencing. After demultiplexing process with CASAVA, following Clontech recommendation, additional 7bp sequencing reads (part of SMART adapter) in the beginning of reads were trimmed prior to mapping.

As shown in Table I, Two samples (59642 and 59643) were prepared for sequencing libraries at different starting amounts. All samples were sequenced at about 10 million reads per samples, with mappable read rate around 74-82%, an acceptable measure in most of RNA-seq studies utilizing FFPE specimens. Of note, sequence read duplication rate decreases when starting material amount is lower (from about 74% to 34%), indicating the reduced RNA diversity at lower starting RNA amount. These findings provide important guidance to ongoing studies toward the overall objective of this project. Specifically, the finding suggest that an input amount of 10ng would be desired in ensuing experiments.

Table I: Summary of RNA-Seq mapping results.

RNA Samples	Sample Name	cDNA synthesis starting amount	Total read (millions)	Mappable reads (percent) Millions (%)	Duplication rates (%)
59462	59462-200	200pg	10.05	7.51(74.7%)	73.66
	59462-2	2ng	11.00	8.60(78.2%)	45.59
	59462-10	10ng	10.86	8.19(75.4%)	34.13
59463	59463-200	200pg	10.22	7.81(76.4%)	74.82
	59463-2	2ng	9.67	8.00(82.7%)	54.15
	59463-10	10ng	11.17	9.17(82.1%)	24.5

Next, we measured gene expression levels using TopHat aligner (version 2.0.8) and HTSeq (version 0.5.4). Sequence read counts were then converted to RPKM by considering transcript length and library size. Genes are considered as expressed genes if their expression level (RPKM) is greater than 1.0. Table II summaries the number of genes detected in these experiments. The results are comparable with published literature suggesting overall good quality of RNA-Seq data when limited amount of FFPE tissues are used.

Table II: Number of genes detected by RNA-Seq.

RNA Samples	Sample Name	# of expressed genes	
		RPKM >1	RPKM > 2
59462	59462-200	8,293	7,375
	59462-2	13,332	12,097
	59462-10	14,699	13,087
59463	59463-200	6,778	5,778
	59463-2	12,056	10,931
	59463-10	14,304	12,691
Merged	59462_all	14,120	12,594
	59463_all	13,446	11,908

A number of key performance characteristics were further evaluated to support the feasibility of using FFPE tissues for RNA-Seq for the specific purpose of comparing low-risk and high-risk prostate cancer. Figure 3 shows the mapping rates for exon, intron, and inter-genic sequences. The data suggest minimal effect of the starting amount of RNA on mapping results. Figure 4 shows the % coverage rate for the 5' and 3' of the genes, supporting uniform coverage. Another important measure is % rRNA depletion. Relevant findings on rRNA depletion as a result of input FFPE RNA amount is shown in Figure 5. Sample number 59463 had better rRNA depletion profile than sample number 59462, possibly reflecting better RNA quality (not shown) in 59462. Figure 6 presents Pearson correlation of top 1000 high expression genes between the two low-input samples and sample with 10ng input. The data suggest low data quality in samples with low RNA input. In Figure 7, we present data on average coverage by gene position for the top 1000 expressed genes. Overall, these standard data quality measures support the conclusion that high quality RNA-Seq can be obtained from FFPE RNA in the nanogram range, on the basis of comparable performance characteristics established in current literature.

### **Year 3:**

During year 3 of the project period, we focused on analysis of clinical specimens suitable for RNA sequencing studies. We summarized our efforts in our 2015 annual report.

1. We identified all surgical cases from 2014 to 2015 (to minimize the age effect).
2. Working with the clinical staff members, we finalized the case selection parameters following definition of “very low-risk” prostate cancer preoperatively in biopsy specimens, as well as definition of “low-risk” and “upgraded” prostate cancer postoperatively in radical prostatectomy specimens.
3. We finalized a list of 61 surgical specimens meeting the criteria for “indolent” prostate cancer, and a list of 33 surgical specimens meeting the criteria for “upgraded” prostate cancer.
4. The finalized list was later expanded to 70 surgical specimens from which RNA was extracted (see below for details).

### **EWOFF period:**

During this period, we prepared a total of 6 sections from each of the 70 cases (total 420 sections). One section from each case was used for pathological diagnosis and circling of small tumor lesions, and the remaining unstained slides were stored for RNA extraction. RNA extraction of the target lesion was performed using the PureLink FFPE RNA isolation kit (Ambion, Thermo Fisher Scientific), following identification of the target lesions according to the H&E stained adjacent sections. We first optimized the proteinase K incubation time using slides made from cell pellets (Figure 8, Appendix). In Figure 8, duplicated extractions (named left and right) were performed on FFPE sections prepared from LNCaP cell pellets following periods of 1-hour, 2-hour, 3-hour, or overnight (O/N) proteinase K incubation, and RNA yield and quality evaluated by Qubit 3 and Bioanalyzer. Slightly improved yield was obtained following O/N incubation (Figure 8). Following



RNA extraction (O/N incubation) from all 70 cases, we performed RNA evaluation using the Bioanalyzer RNA Pico Chip assay for quality analysis. Due to the large file size for all batches of experiments, we present only one batch of experiment to illustrate the variation among these clinical samples (Figure 9, Appendix). In Figure 9, the electropherogram of RNA from 8 cases were presented, showing RNA degradation in all but one of the cases (Figure 9). The remaining batches of experiments showed similar patterns of poor RNA quality across the clinical specimens. Therefore, although selected cases may be feasible for RNA sequencing (as shown in our preliminary studies reported in our 2014 annual report), we have determined that RNA sequencing is unlikely to succeed in a relatively large cohort of clinical specimens from men with very low-risk prostate cancer. We speculated RNA quality may be compromised by the many technical steps involved in the challenging task of small lesion identification, isolation, digestion, RNA purification, subsequent handling. Given the current technology available to us, we determined that comprehensive analysis of this cohort (n=70) is no longer feasible.

**Findings resulting from Task 2: To validate a refined set of genes predictive or indicative of higher-risk disease within a PAS longitudinal cohort (Months 12-36).**

**Summary:** According to our project plan in SOW we planned to carry out studies related to this task during year 2 and year 3 of the project period. Our main research activities related to this Task was reported in our 2014 annual report. Our research activities focused on identification of biopsies. In our 2014 annual report, we also communicated an expected delay in carrying out the research activities related to Task 2, mainly due to a corresponding delay of Task 1. In light of the additional technical hurdles we have experienced for molecular analysis of small volume prostate tumors detailed in Supporting Data for Task 1, the scientific value of further research related to Task 2 is diminished, and Task 2 is no longer pursued.

**Supporting data (reported in 2014 annual report):**

We have identified a total of 1060 biopsies suitable for studies proposed in Aim 2. These biopsies met the NCCN very low risk prostate cancer criteria (stage T1c, and PSA <10ng/m; Gleason score <=6; and no more than 2 cores containing cancer, and <=50% of core involved with cancer; PSA density <0.15ng/ml/g). A subset of them (n=232) represent those from the patients meeting the entry criteria for the active surveillance program but nevertheless reclassified longitudinally.

For the 1060 available research biopsies within our biorepository, diagnostic biopsies, confirmation biopsies and annual monitoring biopsies of follow up patients are all available and factored in the tally. Upon analysis of diagnostic classification distribution of all available biopsies duplicates, there are in total 828 biopsies from 338 cases in the very-low-risk group while there are 232 biopsies from 186 cases in the biopsy progression group. These specimens are more than sufficient for the proposed studies in Aim 2.

**Findings resulting from Task 3: To define somatic DNA copy number alterations and methylation changes when higher-risk disease develops in men undergoing PAS (Months 1-36).**

**Summary:** We have presented our progress on DNA sequencing in our 2013 annual progress report. A delay in executing Task 3 was expected and communicated in our 2014 annual report. Full results related to Task 3 will be reported in the Final Report. Primarily due to the technical difficulty in working with small volume tumors, Task 3 is not pursued beyond what we reported previously.

**Supporting data (reported in our 2013 annual progress report):**

Technical evaluation using the Illumina HiSeq 2000 platform. To evaluate the various technical aspects of deep sequencing, we subjected DNA samples to X chromosome specific exome capture followed by sequencing using the Illumina HiSeq 2000. The short read sequences (50bp) were aligned by BWA aligner (5). The copy number alterations (CNA) were determined by the following sequential steps: First, the sequence depth was calculated by SAMTools (6); Second, the copy number changes were estimated by comparing the sequence depth to that from normal samples through the VARSCAN software (7); Finally, the copy number segmentations were determined by Circular Binary Segmentation (CBS) algorithm (8). The CNA frequencies of representative tumor samples are shown in Figure 10 (Supporting Data), where red color denotes copy number gain and blue for copy number loss. As shown in Figure 10, the gene AR and OPHN1 had high frequencies of copy number gain. The region that contains two CT antigen, CT45A4 and CT45A5, had the largest copy number loss. We would like to note that specimens used in this initial evaluation were not from men qualified for active surveillance. Nevertheless, by reliably identifying CNVs in prostate tumor samples using this platform, we gained essential experience with the platform that allowed us to conduct full evaluation of the technical feasibility for the proposed studies.

**Key Research Accomplishments**

1. Identified sufficient number of surgical cases from men meeting the entry criteria for active surveillance.
2. Established that high quality nucleic acid molecules may be extracted from select small tumor lesions present in target specimens from men meeting the entry criteria for active surveillance.
3. Optimized the technical steps involved in deep sequencing.
4. Established that high quality RNA sequencing data can be generated from limited amount of input RNA isolated from FFPE specimens, for the specific comparison of low-risk and high-risk prostate cancer.
5. Identified sufficient number of biopsy cases and sections.

6. Finalized and acquired a set of clinical specimens suitable for RNA sequencing following time-consuming efforts, and evaluated RNA quality and yield from a cohort of 70 cases.

## **Reportable Outcomes**

Manuscripts: None at this time.

Presentations: None at this time.

Grant Applications:

**Title:** Reducing Prostate Cancer Overdiagnosis and Overtreatment (NIH P01, PI: Pienta)

**Supporting Agency:** NIH/NCI

**Performance Period:** 7/1/2015 - 6/30/2020

**Level of Funding:** \$310,000

**Role:** Project Lead, Project 2 (resubmission)

**Status:** not funded

## **Conclusion**

We conclude that many technical hurdles encountered in molecular analysis of very-low risk prostate cancer pathological specimens may be addressed through carefully planned technical evaluation strategies. We also conclude that while molecular studies of a larger cohort of specimens remains feasible, an essential requirement is sample quality control and meticulous processing. In the absence of a standardized, reproducible, and well-established procedure, large-scale molecular studies focusing on very small tumor lesions may not be feasible given the current technological setting. Clinical processing of the specimens is currently beyond the control of laboratory scientists, emphasizing the need for effective communication and establishment of a clinical workflow that is compatible with future efforts in high-throughput molecular profiling of small pathological lesions.

## **References**

1. Dunn TA, Fedor H, Isaacs WB, De Marzo AM, Luo J. Genome-wide expression analysis of recently processed formalin-fixed paraffin embedded human prostate tissues. *The Prostate*. 2009;69(2):214-8. Epub 2008/11/01. doi: 10.1002/pros.20863. PubMed PMID: 18973134; PubMed Central PMCID: PMC2612089.
2. Dunn TA, Fedor HL, De Marzo AM, Luo J. Molecular profiling of indolent human prostate cancer: tackling technical challenges to achieve high-fidelity genome-wide data. *Asian journal of andrology*. 2012;14(3):385-92. Epub 2012/02/07. doi:

- 10.1038/aja.2011.147. PubMed PMID: 22306912; PubMed Central PMCID: PMC3433951.
3. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, Gnirke A, Pochet N, Regev A, Levin JZ. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature methods*. 2013;10(7):623-9. Epub 2013/05/21. doi: 10.1038/nmeth.2483. PubMed PMID: 23685885; PubMed Central PMCID: PMC3821180.
  4. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC genomics*. 2014;15:419. Epub 2014/06/04. doi: 10.1186/1471-2164-15-419. PubMed PMID: 24888378; PubMed Central PMCID: PMC4070569.
  5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PubMed Central PMCID: PMC2705234.
  6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352 btp352 [pii]. PubMed PMID: 19505943; PubMed Central PMCID: PMC2723002.
  7. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009;25(17):2283-5. Epub 2009/06/23. doi: 10.1093/bioinformatics/btp373 btp373 [pii]. PubMed PMID: 19542151; PubMed Central PMCID: PMC2734323.
  8. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004;5(4):557-72. Epub 2004/10/12. doi: 5/4/557 [pii] 10.1093/biostatistics/kxh008. PubMed PMID: 15475419.

## **Appendices**

### **Supporting Data (10 figures and figure legends)**

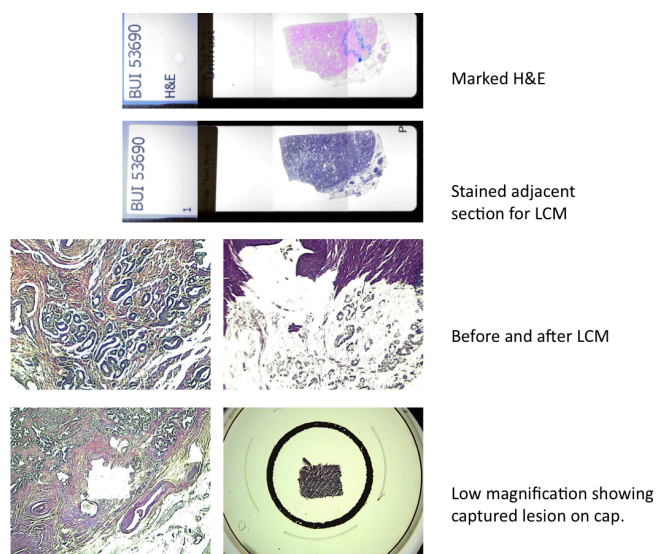


Figure 1: Laser capture microdissection of a marked tumor lesion present in a surgical specimen from a patient meeting the entry criteria for active surveillance.

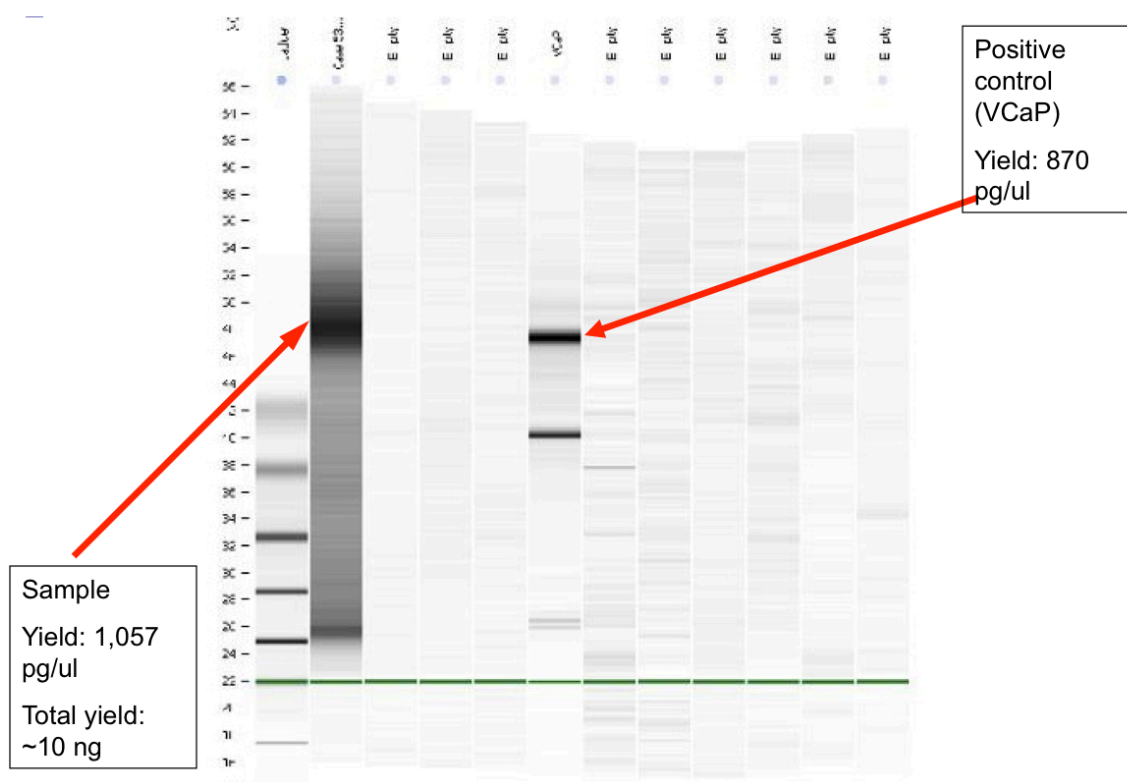


Figure 2: Electropherogram of RNA extracted from laser captured small tumor run on the Agilent Bioanalyzer. The red arrow to the left points to the presence of the 28S rRNA, a

marker of RNA quality sufficient for genome-wide RNA analysis. The total yield from the captured lesion is 10ng, also sufficient for the proposed studies. The second red arrow points to a RNA sample prepared from standard cell lines.

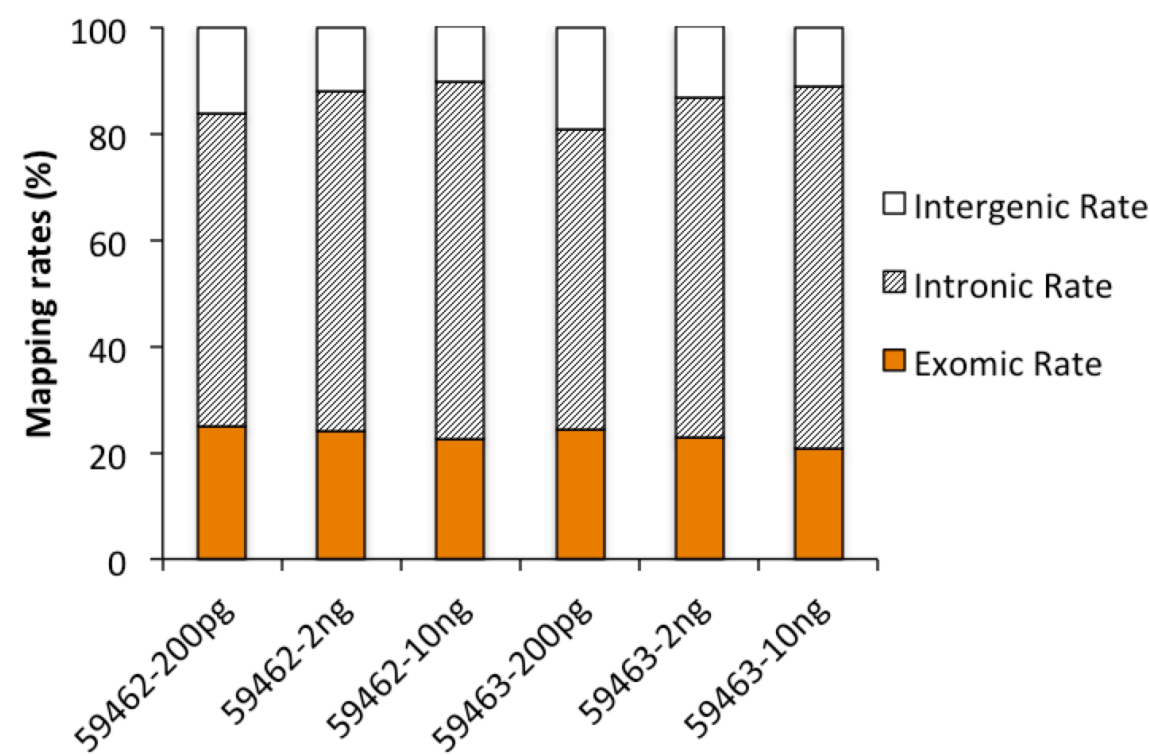


Figure 3: Percentage of sequencing reads mapped to exons, introns, and intergenic regions of the human genome by varying amounts of input FFPE RNA.

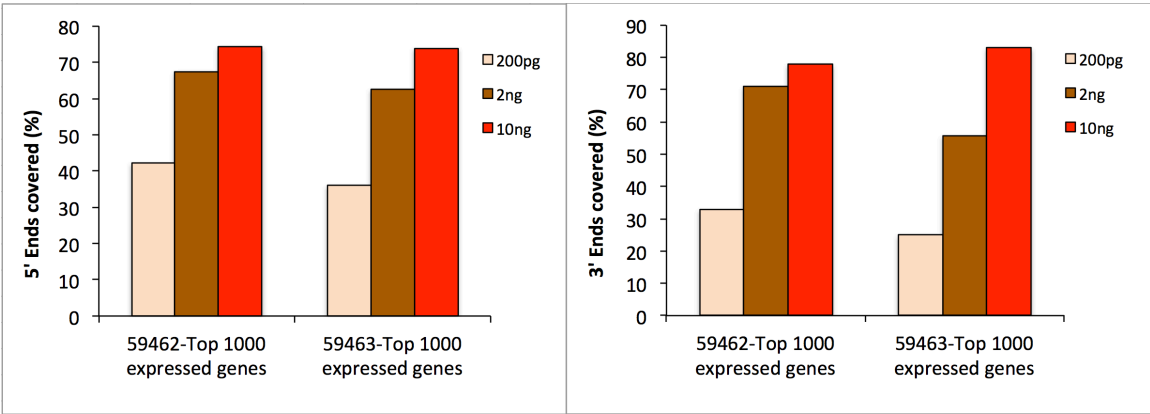
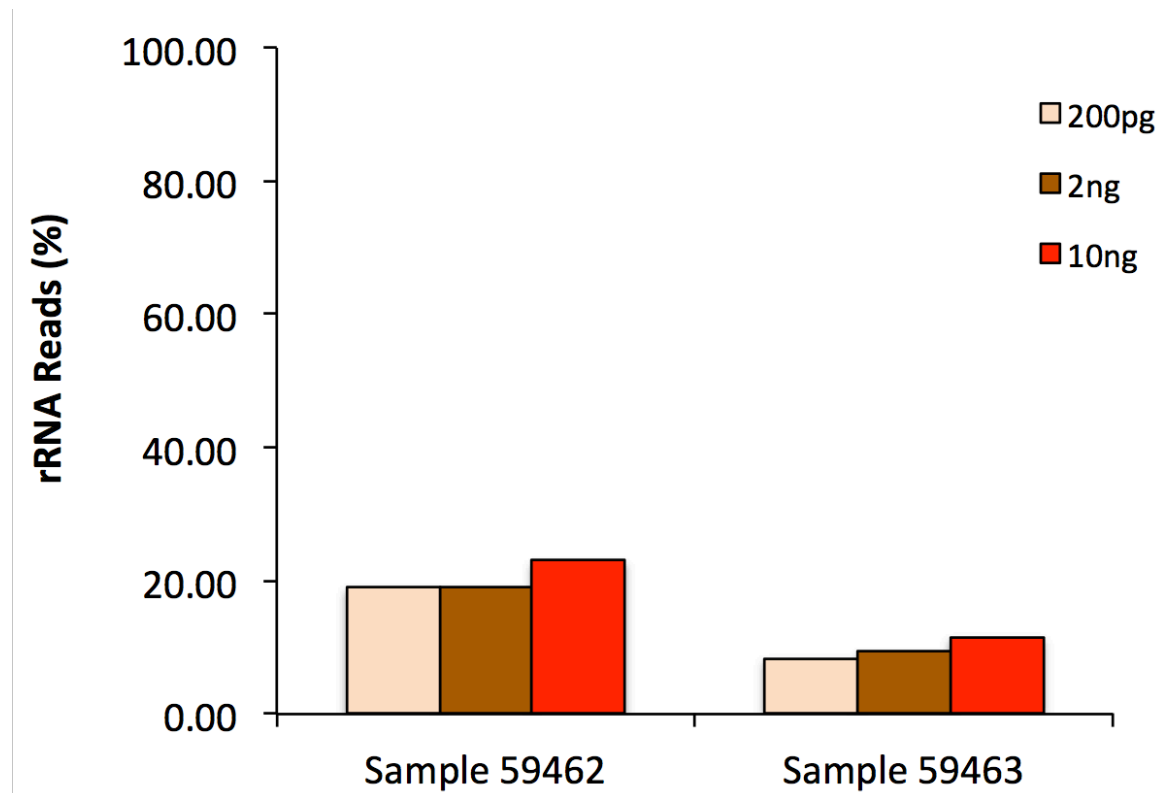


Figure 4: Sequence coverage at the 5' and 3' of the gene transcripts for the top 1000 expressed genes determined by RNA-Seq.



**Figure 5.** Efficiency of rRNA depletion by sample type and varying amounts of input FFPE RNA.

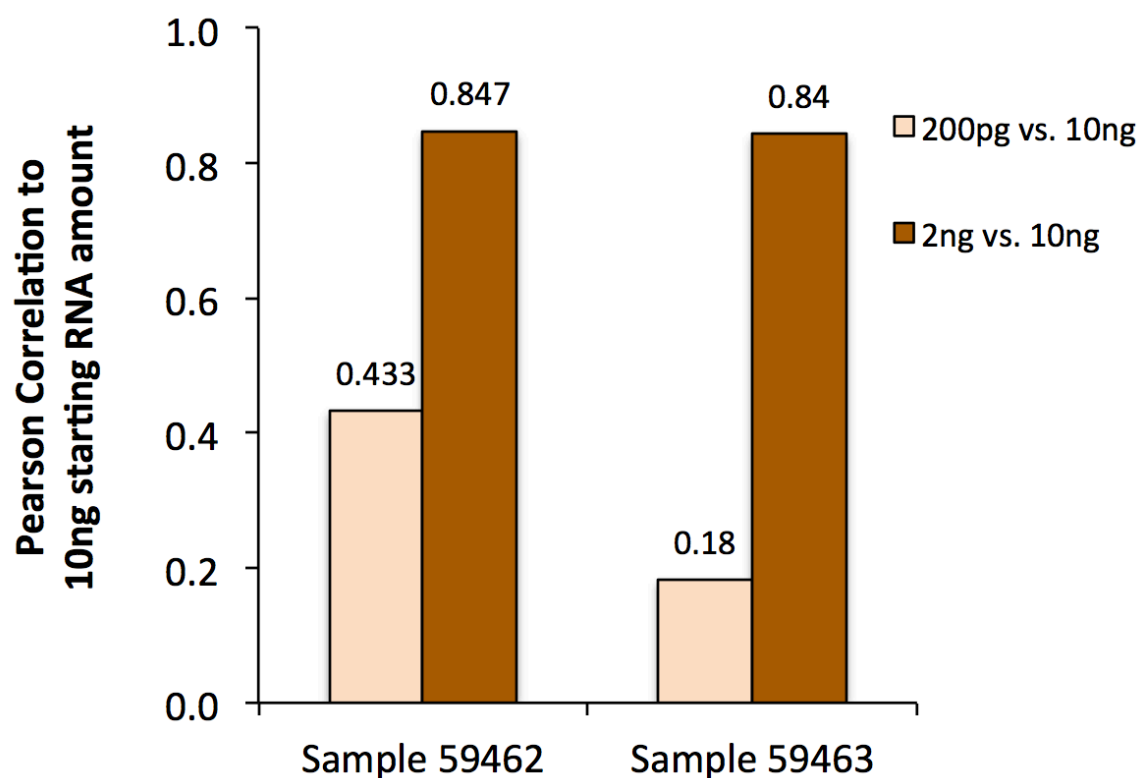
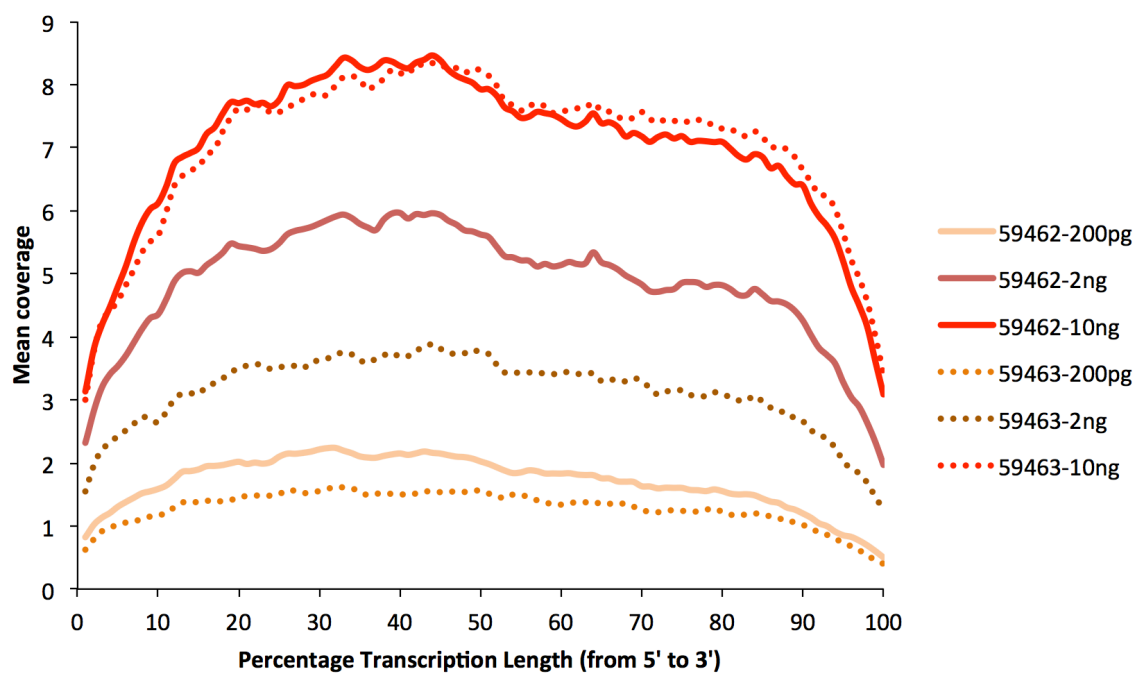
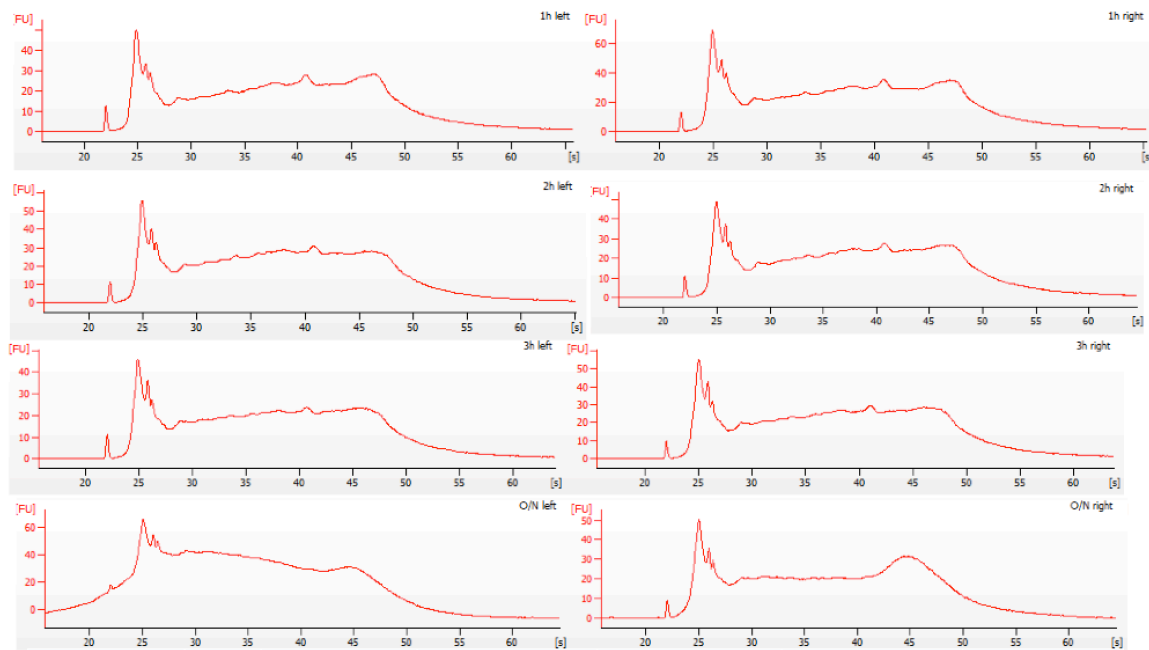


Figure 6: Correlation of transcript abundance between RNA-Seq data derived from lower input RNA (200pg and 2 ng) versus 10ng RNA.



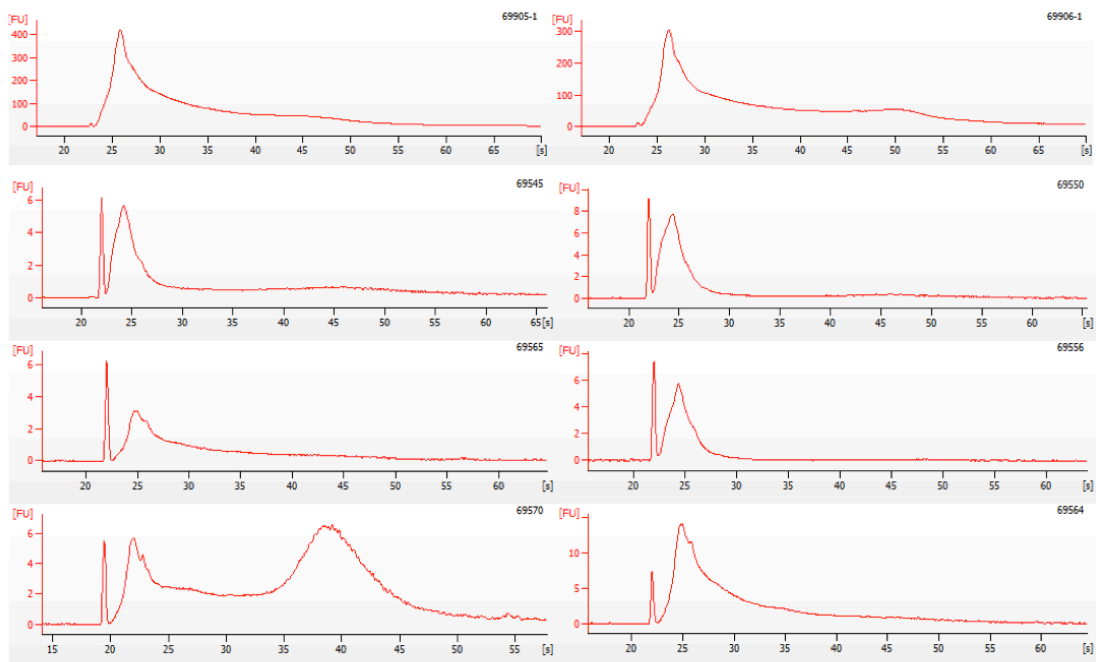


**Figure 7:** Mean coverage plot by position for top 1000 highly expressed genes determined by RNA-Seq.

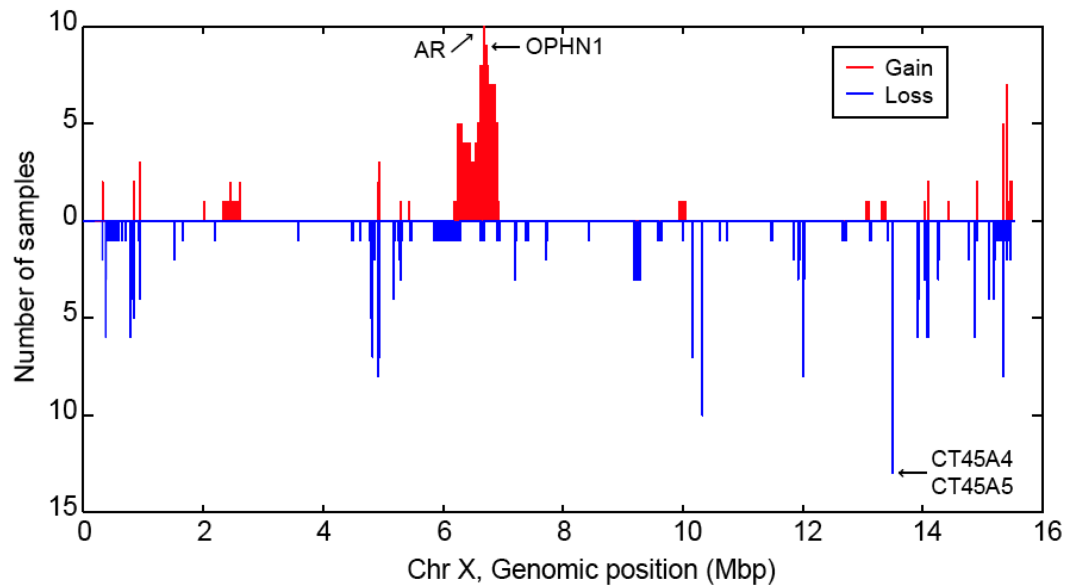


		Sample ID	Qubit 3 [RNA] (ng/ul)	Total yield (ug) (Qubit)	RNA used in RT (ul)	[RNA] from Bioanalyzer (pg/ul)	RIN from Bioanalyzer
7/28/2016	1	1h left	14	0.42	20	10979	2.4
7/28/2016	2	1h right	19.7	0.591	20	14506	2.3
7/28/2016	3	2h left	19.9	0.597	20	12795	2.2
7/28/2016	4	2h right	18.6	0.558	20	11193	2.3
7/28/2016	5	3h left	16.7	0.501	20	10255	2.2
7/28/2016	6	3h right	22.4	0.672	20	12289	2.2
7/28/2016	7	O/N left	20.4	0.612	20	19670	2
7/28/2016	8	O/N right	24.4	0.732	20	11580	N/A

**Figure 8:** Bioanalyzer analysis of RNA samples from LnCaP FFPE slides with different proteinase K incubation time [Each of the four time points has duplicated samples, labeled left or right due to the position on the FFPE slide]



**Figure 9: Bioanalyzer electropherograms of RNA extracted from eight histologically defined very-low risk prostate cancer FFPE slides. Note that with the exception of case #69570 (marked at the top right of each electropherogram), all samples were determined to be of low quality and not suitable for molecular analysis.**



**Figure 10.** DNA copy number alternations from 14 prostate tumor samples, comparing to copy numbers from 2 normal prostate tissues. Red color denotes copy number gain and blue for copy number loss. The gene *AR* and *OPHN1* had high frequencies, 10 and 9 out of 14 samples, respectively, of copy number gain. The region that contains two CT antigen, *CT45A4* and *CT45A5*, had the largest copy number loss.